PREDICTION OF PLAN ADAPTATION IN HEAD AND NECK CANCER PROTON THERAPY USING CLINICAL, RADIOGRAPHIC, AND DOSIMETRIC FEATURES

A Dissertation
Presented to
The Academic Faculty

by

Duncan Bohannon

In Partial Fulfillment
of the Requirements for the Degree
Masters of Science in the
School of Mechanical Engineering
Department of Nuclear & Radiological Engineering and Medical Physics

Georgia Institute of Technology
August 2021

PREDICTION OF PLAN ADAPTATION IN HEAD AND NECK CANCER PROTON THERAPY USING CLINICAL, RADIOGRAPHIC, AND DOSIMETRIC FEATURES

Approved by:

Dr. Jun Zhou Advisor
Department of Radiation Oncology
Emory University Hospital

Dr. Chris Wang
School of Mechanical Engineering
Georgia Institute of Technology

Dr. Steven Biegalski
School of Mechanical Engineering
Georgia Institute of Technology

Date Approved: July 27, 2021

ACKNOWLEDGEMENTS

I would like to thank my advisor Dr. Zhou for his endless support during this project. Without him, this thesis would not be possible. I would also like to thank Dr. Wang for recruiting me into the medical physics program and being an invaluable teacher and mentor. I want to thank my classmates William LePain, Clara Glassman, and John Demoor for their contributions to my knowledge of medical physics through our courses together. I want to thank my parents Jeanne Law-Bohannon and Chuck Bohannon for their immeasurable support in my life and education. Lastly, I want to thank my friends Fatima Boumahdi and Maliha Hasan for their constant encouragement during the writing of this thesis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

H&N – Head and Neck

IMRT – Intensity Modulated Radiation Therapy

IMPT – Intensity Modulated Proton Therapy

OAR – Organ at Risk

EPTC – Emory Proton Therapy Center

WET – Water Equivalent Thickness

RSP – Relative Stopping Power

pCT – Planning Computed Tomography

CT – Computed Tomography

PBS – Pencil Beam Scanning

SOBP – Spread-Out Bragg Peak

CBCT – Cone Beam CT

QACT – Quality Assurance CT

CTV – Clinical Target Volume

SIB – Simultaneous Integrated Boost

TPS – Treatment Planning System

BHI – Beam Heterogeneity Index

DVH – Dose Volume Histogram

MLP – Multilayer Perceptron

TP – True Positive

TN – True Negative

FP – False Positive

FN – False Negative

ROC - Receiver Operating Characteristic

SUMMARY

Proton therapy has great advantage over photon therapy in head and neck cancer radiotherapy due to the absence of exit dose from the proton beams. However, proton therapy is much more sensitive to anatomical changes and setup uncertainties, which results in up to 40% re-planned rates in head and neck cancers (H&N). Using clinical data from over one hundred H&N patients treated with proton therapy, this project aims to create a neural network which will be trained with patients' clinical, radiographic, and dosimetric information, and will be able to evaluate plan quality and predict the probability of plan adaption. This will be a great tool to guide clinical practice in proton therapy for not only head and neck cancer treatments, but also other treatment sites that suffer frequent plan adaptations.

CHAPTER 1. INTRODUCTION

1.1 Motivation

H&N cancers are a diverse group of tumors with many pathologies and etiologies. They cause an estimated 330,000 deaths worldwide (1). H&N cancers account for 3% of all cancers and have various subsites that include the salivary glands, oral cavity, and pharynx (2,3). Though there is great diversity within H&N cancers, they are all very difficult to treat because of tumor proximity to critical organs.

Radiation therapy is used in conjunction with surgery and chemotherapy to combat H&N tumors, even being able to treat early-stage tumors on its own. Intensity modulated radiation therapy (IMRT), a type of photon therapy, is a common choice of treatment for H&N cancers. Intensity modulated proton therapy (IMPT) has become an increasingly popular choice for H&N cancer treatment because of their highly accurate dose local deposition and steep dose gradients.

In cases where anatomy changes or setup uncertainties are too severe, the original treatment plan is no longer suitable, and adaptive therapy is necessary. Adaptive therapy is a concept that was originally introduced for photon therapy over 20 years ago (4). Studies have shown that photon adaptive therapy has increased the sparing of organs at risk (OARs) in H&N sites (5,6).

Adaptive therapy is much more frequently needed for proton treatment. It has been shown that adaptive therapy is necessary for many proton sites (7-11). At the Emory Proton Therapy Center (EPTC), 30-40% of H&N cases are re-planned at least once

during their course of treatments. It takes up to 5 days for a re-plan adaptation to be used for treatment. H&N re-plans make up most re-plans across all sites at our institution for all the treatments in 2020, as shown in Figure 1.

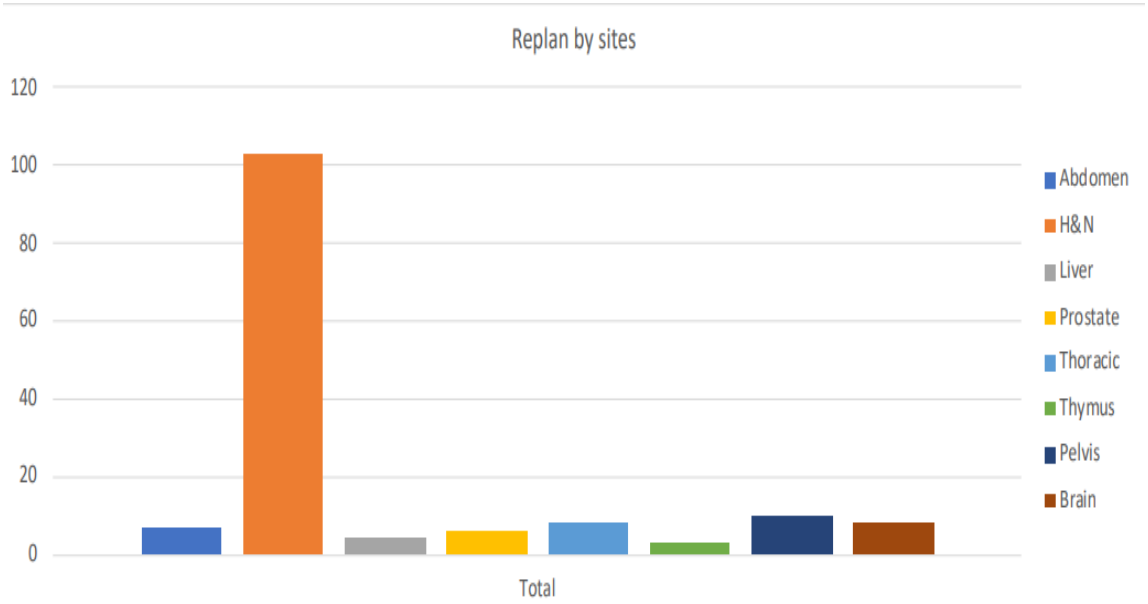

Figure 1 Re-plan cases across all treatment sites at EPTC in 2020

1.2 Photon vs. Proton Therapy

Proton therapy has several advantages over conventional photon therapy. In photon therapy, the dose-vs-depth curve continuously and exponentially decreases through the patient, which causes damage to healthy tissue distal to the target. In proton therapy, however, there is a sharp dose fall-off and the maximum dose is at the end of their range (12). The linear energy transfer of protons remains mostly constant until they start to significantly slow down, leading to a rapid deposition of energy at the end of their particle track known as the Bragg peak (13). The differences between the dose deposition of photons and protons are shown in Figure 2.

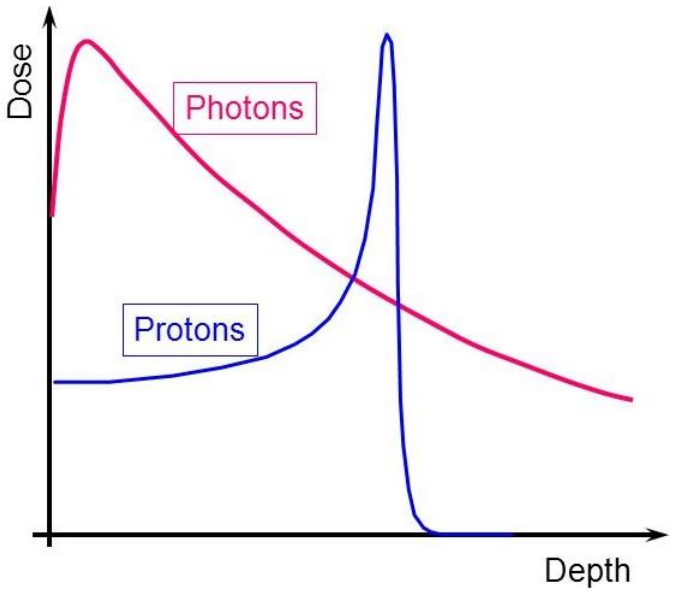

Figure 2 - Depth Dose Curves for 6MV Photons and 120MeV Protons

The position of the Bragg peak is a function of the proton energy and the water equivalent thickness (WET). WET is calculated from the stopping power relative to water (RSP). RSP is derived from the planning computed tomography image (pCT) using the well-established stoichiometric calibration method that converts HU to density to RSP (14). However due to non-unique density to RSP conversion and many other uncertainties in computed tomography (CT) imaging, +/-3.5% range uncertainties are adopted to account this process, with 95% confidence level. At EPTC, protons are delivered using the pencil beam scanning (PBS) technique, in which proton spots are scanned across the target. Various proton energies with proper weighting are used to create a spread-out Bragg peak (SOBP) to cover the target in the beam direction.

Proton therapy provides tumor coverage while offering more sparing for critical organs than photon therapy. In sites where the proximity of critical structures is a concern, protons are the preferred choice. For these reasons, IMPT has become common in the

treatment of head and neck cancer. However, IMPT is much more sensitive to inter-fractional and intra-fractional changes.

1.3 Inter-Fractional and Intra-Fractional Changes in H&N Radiotherapy

Intra-fraction changes (organ motion, patient motion, breathing) and inter-fraction changes (position uncertainties and anatomic changes) can significantly alter the dose distribution near the targeted tumor boundary in IMPT. H&N sites typically have minimum intra-fractional changes. Inter-fractional changes are much more prevalent, and there are several ways these are mitigating.

Inter-fraction changes are lessened with immobilization devices such as vacuum casts as well as taking regular images to monitor the patient's anatomy. H&N cancer patients undergo several types of anatomical changes through the course of their treatments. These include target shrinkage, weight loss/gain, cavity filling changes, etc. (15-17). Several models have been designed to predict the weight loss of the patient (18-20).

1.3.1 Anatomy Changes

During treatment course, anatomy changes are monitored with cone beam CTs (CBCTs) and quality assurance CTs (QACTs). An example of the impact of anatomy changes on proton therapy plans is shown in Figure 3.

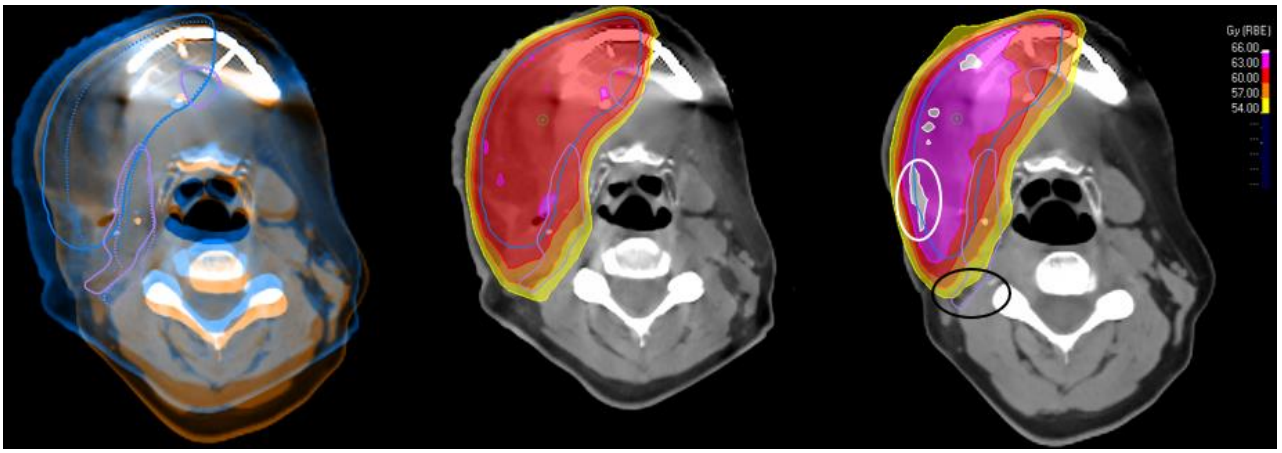

Figure 3 – Image overlay between the pCT and QACT (left, blue: extra tissue in the pCT, brown: extra tissue in QACT), dose distribution in pCT (center), and QACT (right). White circle emphasizes > 110% dose increase, while the black circle includes the under coverage for the CTV54Gy

The left image shows the pCT with solid contours (blue) overlayed with a QACT taken a month later with dashed contours (orange). The blue and purple contours represent the high dose level and intermediate dose level clinical target volumes (CTVs), which appear smaller on the QACT. Even in just a month, this patient has undergone large changes, especially the shrinkage of the right and right anterior face. Also, the posterior neck tissue does not fit the neck rest cushion at the treatment site. The middle image shows the initial plan on the pCT, which has adequate coverage of both CTVs. The right image shows the initial plan on the QACT, which has changed significantly. In the white circle, there is a hot spot and there is a loss of coverage in the black circle. Hot spots and losses of coverage come from the shift in the Bragg peak positions of proton spots, which has to be addressed with plan adaptation.

1.4 Neural Networks and Their Applications in Radiotherapy

Neural networks are computational models made of multiple layers of data processing. Given adequate training data, these models can learn the relationships between independent and dependent variables through modifying model parameters using a back projection algorithm. With enough layers, neural networks can be sensitive to small changes and unsensitive to large variations in the data. They can achieve results comparable to traditional statistical models, and they are used for various types of problems such as pattern recognition and classification (21).

1.4.1 Neural Network Background

The structure of a neural network is shown in Figure 4.

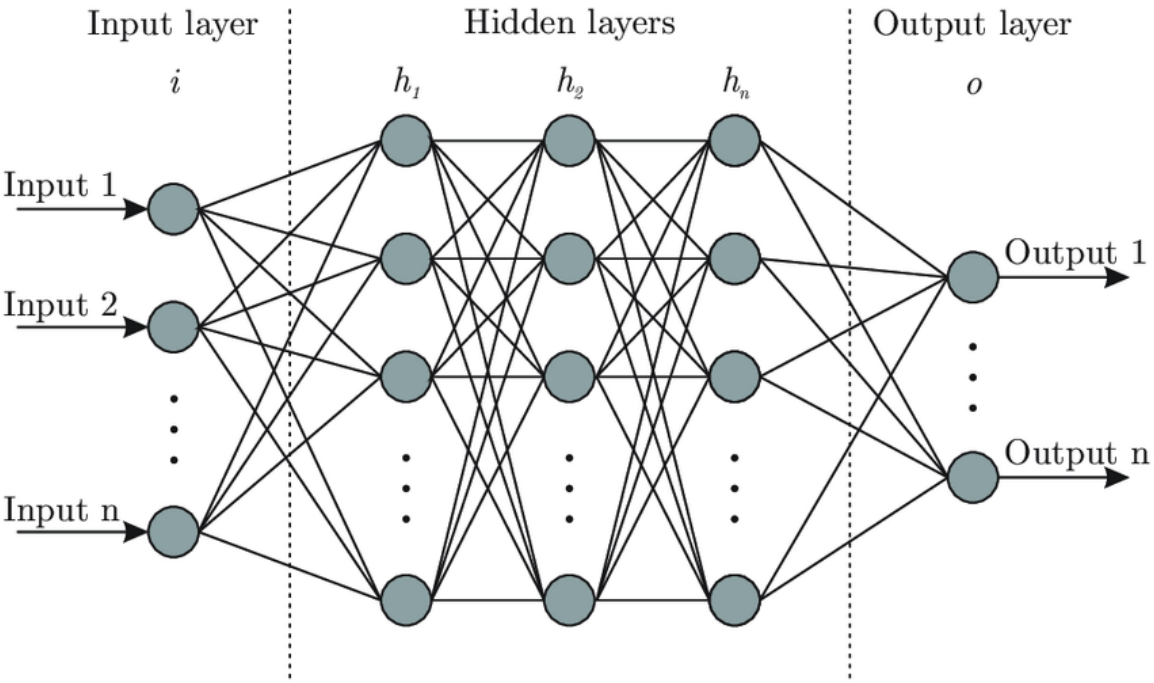

Figure 4 - Example Neural Network Structure

Each layer consists of several nodes. The input layer takes in the independent variables. The hidden layer is used for data processing, and the output layer is the prediction the

model makes. In classification problems, which is what neural networks are used for in this project, the output layer produces the categorical class label.

Each connection between the nodes has a weight w_{ij}. The input of node j in the first hidden layer, I_j, is calculated from the sum of the product of the weights and their corresponding input variables, x_i shown in Equation 1.

$$I_j = \sum_{i=0}^{n} w_{ij} x_i \tag{1}$$

This input is then put into an activation function f. Activation functions in hidden layers are used to help the neural network learn complicated relationships between the data. The output of the node j, O_j is shown in Equation 2.

$$O_j = f(I_j) \tag{2}$$

The choice of activation functions in the output layer depends on the desired output. For classification problems, the softmax function, defined in Equation 3, is used.

$$\sigma(\vec{o}) = \frac{e^{\vec{o}}}{\sum_{m=1}^{n} e^{o_m}} \tag{3}$$

Where $\vec{o}$ is the vector of outputs and n is the number of outputs. Softmax transforms the outputs such that their total equals 1. For multiclass classification, these transformed outputs represent the probability of each output. The neural network then choses the output with the highest probability.

1.4.2 Neural Network Training

Neural networks “learn” from examples given to them. In order to effectively train them, the training data must be representative of the situation being modeled.

The values of the weights in a neural network are given an initial value. Using the inputs taken from the training data, it produces an output. The error between this predicted output and expected output is then calculated. The error is given to a backprojection algorithm, which adjusts the values of the weights. Every training example adjusts the weights, and, eventually, the neural network “learns” the relationship between the data.

The more data given to a neural network, the more powerful the predictive power. However, neural networks can often be overfitted to the data, making them useless in cases not found in the data set.

The choice of backprojection algorithm, scaling of the independent variables, activation functions, and network structure are all important considerations when designing a neural network.

1.4.3 Limitations and Variations

When models have several layers, training a neural network can take a significantly long time. They also require amounts of data in order to be adequately trained. They can also become over correlated to the training data, making the model useless for cases outside the training data. To tell when a neural network is overfitted, training data is split into a training set and a testing set. The testing set is used to see how well the model fits new data, and it validates the predictive power of the neural network.

Because the split of the data is random, neural networks trained with the same data will produce different results. It is necessary to train several different ones to find the best option.

1.4.4 Using Neural Networks for Re-Plan Prediction

Neural networks have been applied to radiotherapy for dosimetric prediction (22-24). This project proposes that a neural network can be designed and used for re-plan prediction. Its training data would come from clinical features and dosimetric parameters from the initial plans of patients treated at our institution and output whether new patients will need a re-plan. This pre-treatment neural network could then be used by dosimetrist and physicists to evaluate the plan prior to treatment and improve the plan when necessary, potentially reducing the re-plan rates.

A separate neural network could be used to monitor the re-plan chance during treatment. This on-line prediction model could be designed by including the changes in WET as an input.

With these neural networks, this project aims to reduce re-plans caused by anatomical changes in H&N cancer.

CHAPTER 2. METHODOLOGY

2.1 Data Collection

The data used to train the prediction model were gathered from 229 H&N patients treated at our institution. Patients spanned 24 different H&N sites, and the most common were the oropharynx (40 cases), lip and oral cavity (17 cases), and nasal cavity/sinuses (16 cases). 40% of patients were re-planned.

2.1.1 Treatment Planning Considerations

At our institution, a 5-beam arrangement is used to treat bilateral H&N patients, which make up 61% of our data. A 3-beam arrangement is used to treat unilateral H&N patients, which make up 39% of our data. These arrangements are shown in Figure 5. Any portion of the target is typically covered by at least two beams, which contribute to the total dose.

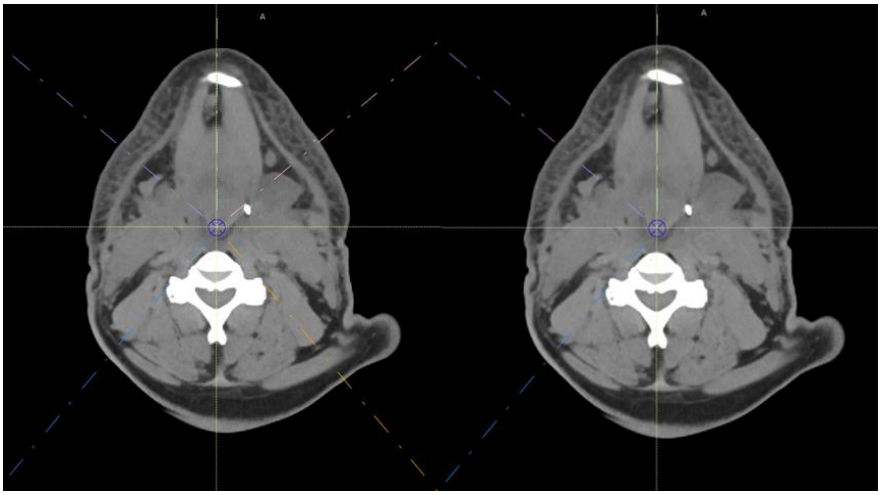

Figure 5 – Typical beam arrangements in H&N proton therapy. Bilateral 5 beam arrangement (left). Unilateral 3 beam arrangement (right).

Most patients are treated with up to 3 dose levels. The highest level is used for the treatment of the gross tumor site or surgical beds. The other dose levels are used to treat lymph nodes, with the higher dose level typically being used on the lymph node closer to the gross tumor site. The number of fractions and dose levels vary between patients, but typical ranges for these levels are 66-70 Gy, 57-61 Gy, and 53-55 Gy treated for 33-35 fractions. All 3 levels are delivered with simultaneous integrated boost (SIB). The coverage goal for each level is D98 = 100%.

Critical structures in HN plans include the spinal cord, brain stem, oral cavity, parotid glands, and optic chiasm. The constraints for these structures at our institution are shown in Table 1.

Table 1 - Critical Structures and Their Constraints in H&N Radiotherapy

Critical Structures	Constraints	
Spinal Cord	<45 Gy at .03cc	
Brain Stem	<54 Gy at .03cc	<60 Gy at .03cc
Oral Cavity	Mean dose < 12 Gy	Mean dose < 35 Gy
Parotid Glands	Mean dose < 26 Gy	< 50% volume at 30 Gy
Optic Chiasm	<54 Gy at .03cc	

A ±3.5% range uncertainty combined with ±3 mm setup shifts on 3 coordinate directions (superior/inferior, anterior/posterior, left/right) was used in robust optimization. For robust evaluation, robustness scenarios consisted of 12 scenarios with setup shifts with range uncertainty and 2 scenarios with just the range uncertainty.

2.1.2 Clinical Features

15 clinical features were manually gathered from the patients using our clinical database. These features are shown in Table 2.

Table 2 – Clinical Features Obtained from H&N Patients

Clinical Features			
Sex	Age	Re-plan (Y/N)	Reason for Re-plan
Site	Stage	TNM	p16 Value
Had Surgery (Y/N)	Surgery Margins	Surgery Date	Planner
pCT and QACT simulation date	Had concurrent chemotherapy (Y/N)		Type of therapy (definitive/palliative)

2.1.3 Dosimetric Features

Various dosimetric features were obtained from patients in the treatment planning system (TPS) Raystation 9A (RaySearch Laboratory, Stockholm, Sweden). These include beam dose heterogeneity index (BHI) and several robustness features. These will be discussed in the next sections. Dosimetric features were chosen to represent plan quality and were gathered from the initial plan and robust evaluation plans. A Python script running in the TPS was made to obtain the data.

2.1.4 Beam Dose Heterogeneity Index (BHI)

There are several ways to define heterogeneity. In this project, BHI was defined as the ratio of the max beam dose to its expected contribution, which was assumed to be half the prescription dose. This is shown in Equation 4.

$$BHI = \frac{D_{max}}{D_{pre}/2} \tag{4}$$

Where D_{max} and D_{pre} are the max dose and prescription dose, respectively.The dose at several small volumes (.03cc, .1cc, .2cc, .5cc) was used to represent the max dose and

dose heterogeneity. The BHI was calculated for every beam. This was used to calculate the max BHI and mean BHI.

2.1.5 Robustness Features

Plan robustness was evaluated with several features gathered from the highest dose level CTV.

The shoulder of the robust evaluation dose volume histogram (DVH) of the CTVs is shown in Figure 5. The nominal plan is shown in the dashed line, and the robustness scenarios are shown in the dotted lines. Several robustness features were obtained from the shoulder of the dose-volume histogram used for robust evaluation. Success rate was defined as the fraction of robustness scenarios where V100 was greater than 95%. The mean V100 CTV and its deviation from the nominal plan, V100 change, defined in Equation 5, were also gathered.

$$Mean\ V100\ Change = \frac{1}{n}\sum_{i=1}^{n}|V100_{nominal} - V100_i| \quad (5)$$

Where n is the number of robustness scenarios.

The last robustness feature, mean max dose change, defined in Equation 6, was defined as the max dose's average deviation from the nominal plan over all the scenarios.

$$Mean\ Max\ Dose\ Change = \frac{1}{n}\sum_{i=1}^{n}|Max\ Dose_{nominal} - Max\ Dose_i| \quad (6)$$

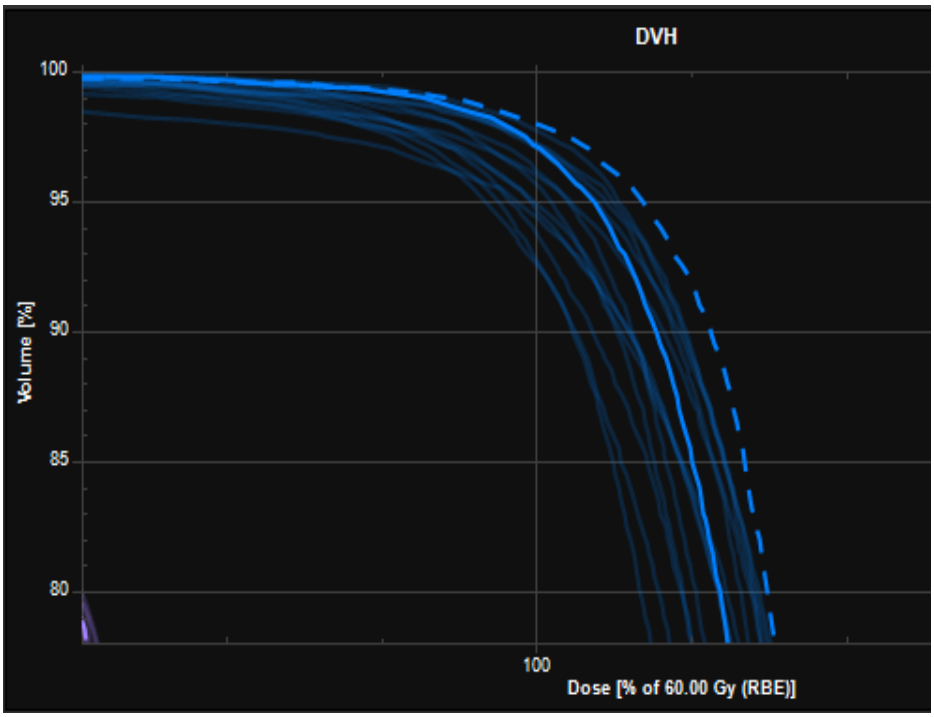

Figure 6 - Shoulder of CTV DVH in robust evaluation with nominal plan (dashed) and robustness scenarios (solid).

2.1.6 Water Equivalent Thickness Calculation

Anatomical changes are a common cause of re-plans. To measure the anatomical changes during treatment course, WET is calculated. WET is the most direct measure of anatomy changes. Figure 7 shows how WET is calculated with a python script in the TPS. Using the image in Figure 7a, the script uses a table to convert HU to RSP. Then, for each beam, the image is rotated by the gantry angle. This moves the image into beam eye's view (BEV).

A BEV image for a 45° beam is shown in Figure 7c. To represent the beam path through the image, WET was integrated along the columns of the BEV image. This shown in Figure 7d. The integrated image was then unrotated as shown in Figure 7e.

Once the unrotated WET CTs for every beam were generated, the WET of the optimization structure corresponding to each beam were found. These structures represent the portion of the CTV that each beam treats. For each beam, its WET was the average WET of all the points on this contour.

The WET script was applied to the pCT and QACTs. The WET of the same beam specific targets in pCT and the differences with the QACT WETs were calculated. This represented patient changes in the beam path between the simulation and the QACT.

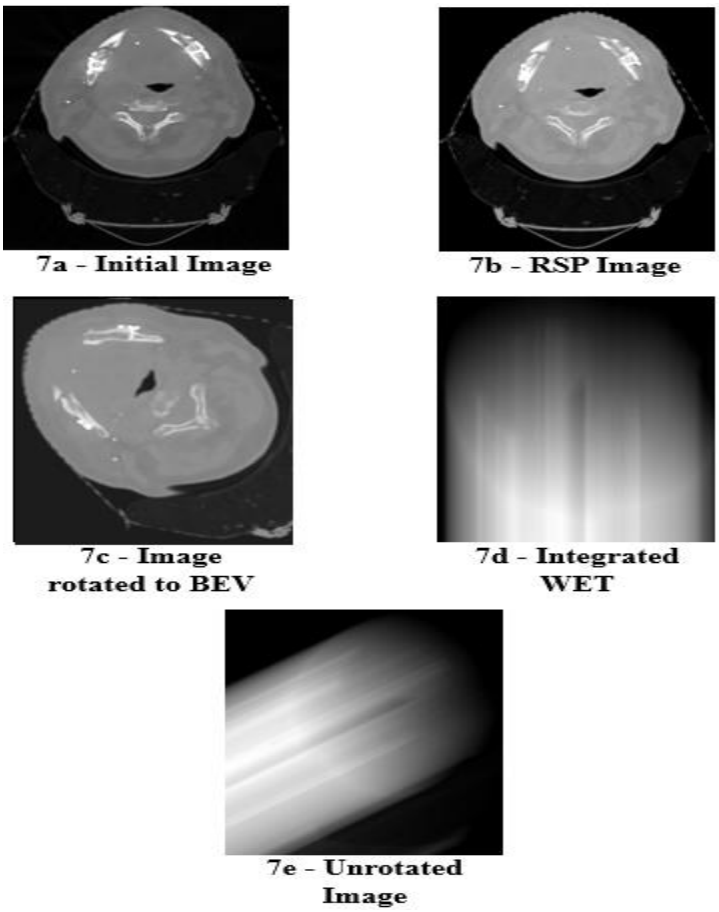

Figure 7 - WET Script Workflow

2.2 Statistical Analysis

IBM SPSS v27 was used to perform statistical analysis on the data. Statistical significance was found to see which features were related to re-plans. To determine the significance, several statistical tests were used. The choice of test depended on the distributions of the data.

Since most of the data were not normally distributed, nonparametric tests were used. All these data were independently sampled, so a Mann-Whitney U test was used. For categorical data, a chi-square test was used to find significance.

2.3 Neural Network Design

2.3.1 Network Structure

The neural networks used for re-plan prediction were designed in IBM SPSS v27 software using its multi-layer perceptron (MLP) function. For both the pre-treatment and on-line models, the structure consisted of an input layer, two hidden layers of 10 and 8 nodes, respectively, and an output layer. A sigmoid activation function was used for the hidden layers and a softmax activation function was used for the output layer. The design of the on-line prediction model is shown in Figure 5. Input 1 to 20: 1) dosimetrist, 2) tumor site, 3) sex, 4) whether the patient had surgery, 5) whether the patient had chemotherapy, 6) time between surgery and CT simulation, 7) tumor stage, 8) T stage, 9) N stage, 10) age, 11) surgery margins, 12) p16 value, 13) max BHI, 14) mean BHI, 15) mean max dose change, 16) success rate, 17) mean V100, 18) mean V100 change, 19) time between CT simulation and QACT, 20) WET % change between pCT and first

QACT. Output of 0 represents no re-plan, and output 1 represents a re-plan. The pre-treatment model did not include inputs 19 and 20, but it was identical to the on-line model otherwise.

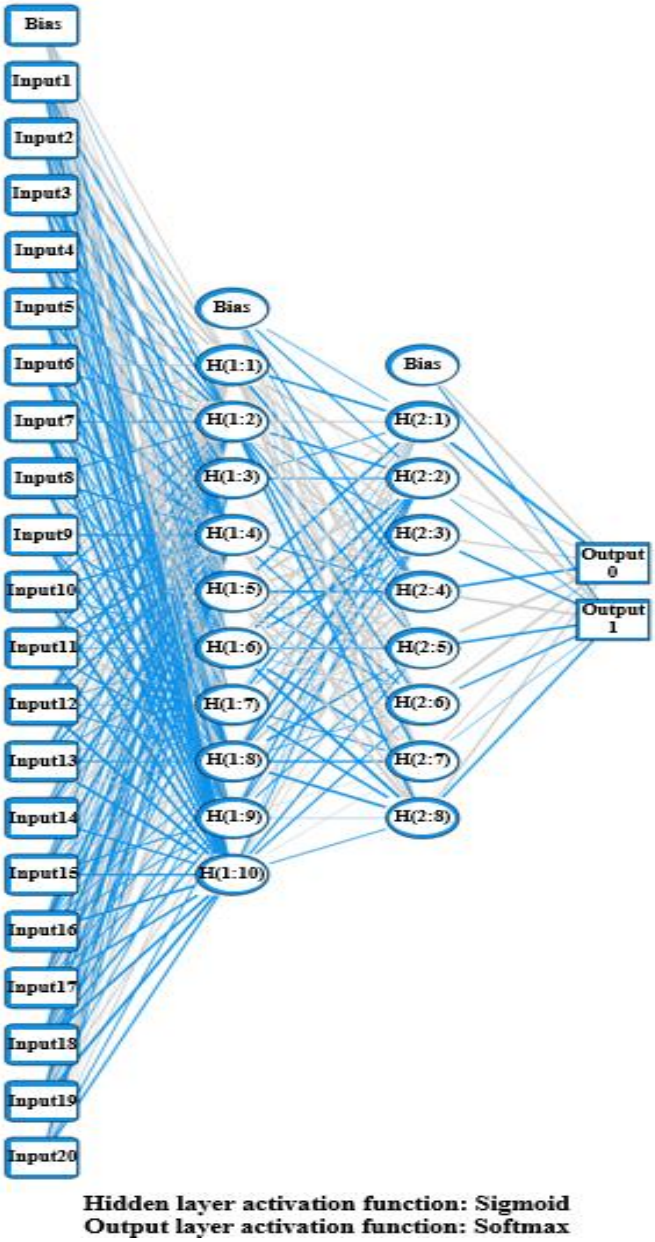

Figure 8 - Multilayer Perceptron Neural Network for Re-Plan Prediction Structure

2.3.2 Training Data

Both models were trained with the dosimetric and clinical features previously mentioned. The data was partitioned into 70% training and 30% testing. Due to patients not passing the export scripts for BHI, robustness, and WET, the training data used is much less than the total number of patients. The initial re-plan prediction model was trained with 53 cases, and the on-line prediction model was trained with 49.

2.3.3 Evaluating Performance

Performance was analysed using a classification table, prediction important index, and receiver operating characteristic (ROC) curves. Classification tables display the true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) of both the training and testing sets. These are used to calculate sensitivity, also called true positive ratio, and specificity, also called the true negative ratio. Sensitivity and specificity are defined in Equations 7 and 8, respectively. To test the impact of the inputs, normalized importance indexes were calculated.

$$Sensitivity = \frac{TP}{TP + FN} \tag{7}$$

$$Specificity = \frac{TN}{TN + FP} \tag{8}$$

CHAPTER 3. RESULTS

3.1 Data Analysis

3.1.1 Numerical Features

All numerical features other than age were normally distributed. Because of this, non-parametric Mann Whitney U-Tests were used for all numerical variables. The null hypothesis was that there was no correlation between the features and re-plans, and the alternative hypothesis was that there was a correlation. A 95% confidence interval was used to determine significance.

Table 3 shows the mean, standard deviation, and p-value for all numerical features. The values shown in Table 3 were calculated after outliers were filtered. These outliers include mean BHI greater than 1.5 or less than .8, mean V100 equal to 0%, and mean max dose change greater than 30%.

Because of the varying passing rates between export scripts, different amounts of data were gathered for different features.

There was no significant difference between the mean/max BHIs with the max dose taken at .03cc, .1cc, .2cc, and .5cc, so the mean/max BHI at .03cc was used to represent BHI.

Time between surgery and pCT, mean/max BHI, and mean max dose change were significant. Age and WET change between pCT and first QACT were weakly significant.

Table 3 – Statistical Parameters of Numerical Features

Feature	Group	Count	Mean	Standard Deviation	P-Value
Age (years)	No Re-Plan Re-Plan	88 46	60.43 66.52	18.90 13.12	p = .054
Time Between Surgery and pCT (days)	No Re-Plan Re-Plan	47 15	37.34 27.60	34.71 9.80	p = .044
Time Between pCT and first QACT (days)	No Re-Plan Re-Plan	72 57	21.85 23.49	6.09 7.19	p = .139
Mean BHI	No Re-Plan Re-Plan	104 61	1.09 1.11	0.10 0.09	p = .020
Max BHI	No Re-Plan Re-Plan	104 61	1.17 1.18	0.16 0.12	p = .020
Success Rate (%)	No Re-Plan Re-Plan	99 53	52.52 52.56	36.20 34.68	p = .785
Mean V100 (%)	No Re-Plan Re-Plan	91 53	94.61 92.61	3.44 13.18	p = .462
Mean V100 Change (%)	No Re-Plan Re-Plan	99 53	3.75 5.61	2.39 13.04	p = .852
Mean Max Dose Change (%)	No Re-Plan Re-Plan	97 52	2.87 3.22	2.56 1.70	p = .009
WET change between pCT and first QACT (%)	No Re-Plan Re-Plan	72 57	2.00 2.40	1.45 2.22	p = .054

3.1.2 Categorical Features

Tables 4 and 5 show the counts of re-plans and non-replans in every category of the statistically significant and weakly significant categorical variables.

Table 4 – Frequency of Re-Plans in Surgery (p = .084)

	Count		
Category	No Re-Plan	Re-Plan	Total
No Surgery	41	30	71
Surgery	47	15	62

Table 5 - Frequency of Re-Plans in Surgery Margins (p = .016)

	Count		
Category	No Re-Plan	Re-Plan	Total
No Margins	46	32	78
Negative Margins	21	8	29
Close Margins	9	4	13
Positive Margins	11	2	13

3.2 Neural Network Performance

3.2.1 Pre-Treatment Model

Table 6 is the classification table for the pre-treatment model. The sensitivity of this model is .750, and the specificity is .774. The model performed worse with the testing data, but still had an average accuracy of 75.7%.

Table 6 - Classification Table for Pre-Treatment Model

		Actual Value	Predicted Value		Total	Overall Accuracy	Total
			No Re-Plan	Re-Plan			
Training Set	Count	No Re-Plan Re-Plan	24 6	7 18	31 24	42	55
	%	No Re-Plan Re-Plan	77.4% 25.0%	22.6% 75.0%	100% 100%	76.4%	100%
Testing Set	Count	No Re-Plan Re-Plan	5 1	2 4	7 5	9	12
	%	No Re-Plan Re-Plan	71.4% 20.0%	28.6% 80.0%	100% 100%	75%	100%

Figure 9 shows the ROC curve for the pre-treatment model. The model performed similarly for both re-plans and non-replans and had an area under the curve of .855.

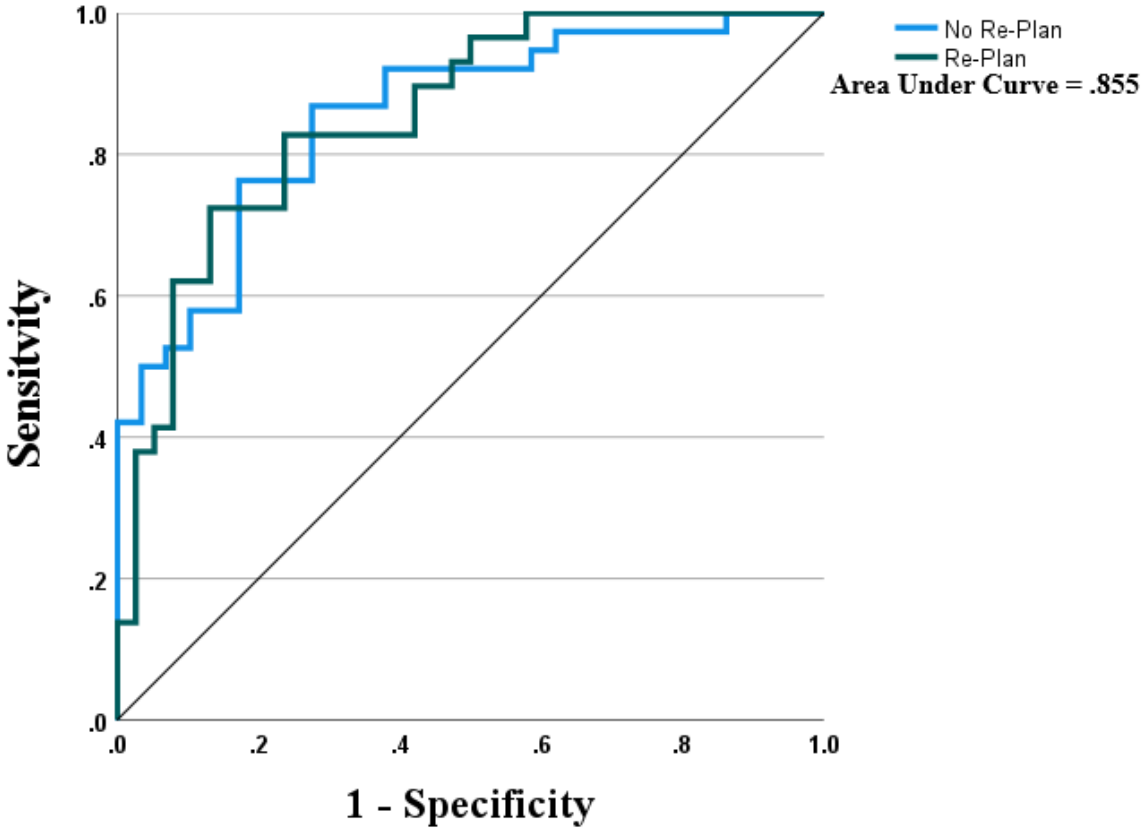

Figure 9 - ROC Curve for Pre-Treatment Model

Figure 10 shows the normalized importance of features in the pre-treatment model. Though importance varies between different models, mean/max BHI, age, mean max dose change, and the time between surgery and pCT were consistently the most important features.

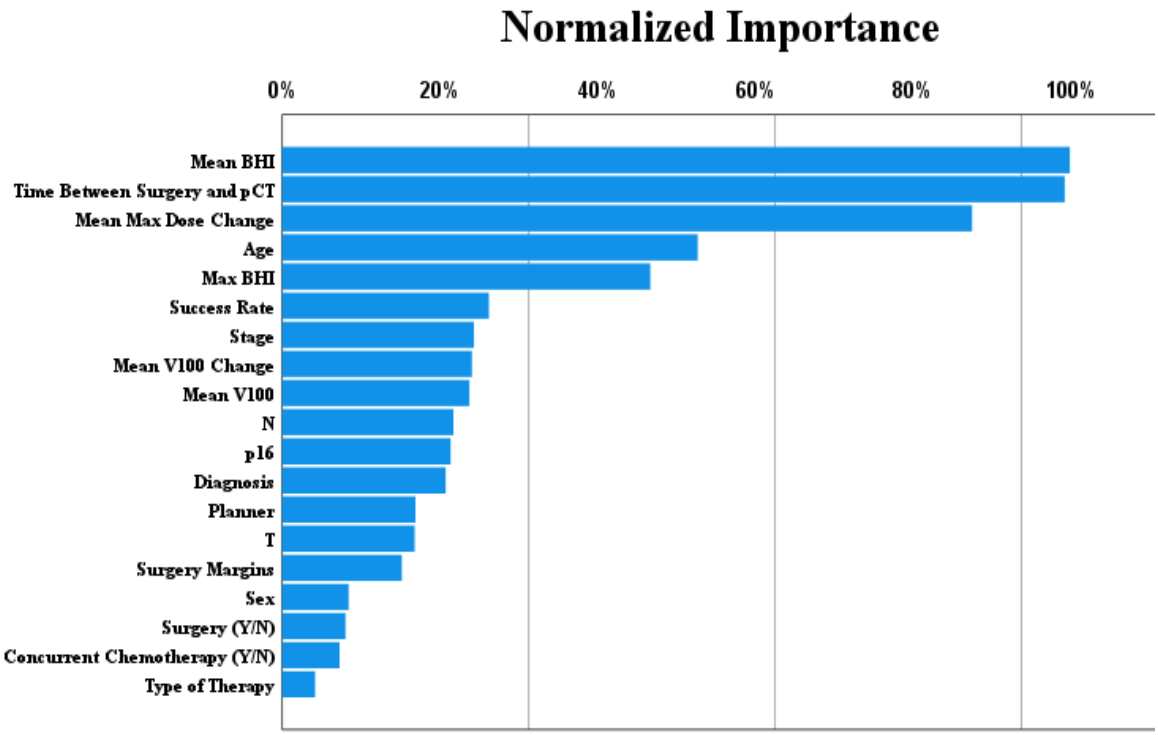

Figure 10 - Normalized importance of features in pre-treatment model. Mean BHI, time between surgery and pCT, and mean max dose change were the most important features.

3.2.2 On-Line Model

Table 7 is the classification table for the on-line model. The sensitivity of this model is .750, and the specificity is .850. The model performed worse with the testing data, but still had an average accuracy of 78.9%.

Table 7 - Classification Table for On-Line Model

		Actual Value	Predicted Value		Total	Overall Accuracy	Total
			No Re-Plan	Re-Plan			
Training Set	Count	No Re-Plan Re-Plan	17 5	3 15	20 20	32	40
	%	No Re-Plan Re-Plan	85.0% 25.0%	15.0% 75.0%	100% 100%	80.0%	100%
Testing Set	Count	No Re-Plan Re-Plan	2 0	2 5	4 5	7	9

Table 7 continued

%	No Re-Plan Re-Plan	50.0% 0	50.0% 100%	100% 100%	77.8%	100%

Figure 11 shows the ROC curve for the on-line model. The model performed similarly for both re-plans and non-replans and had an area under the curve of .885.

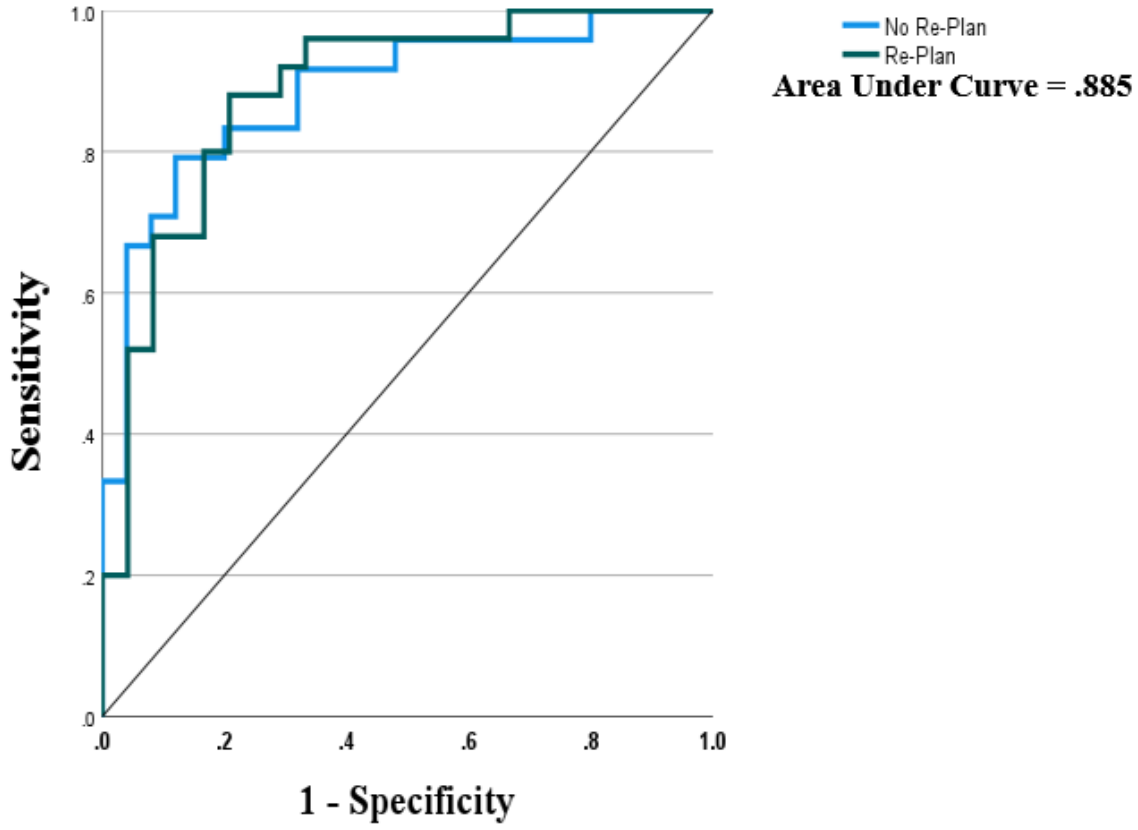

Figure 11 - ROC Curve for On-Line Model

Figure 12 shows the normalized importance of features in the on-line model. Though importance varies between different models, mean/max BHI, age, mean max dose change, and time between surgery and pCT were consistently important features. Though WET change between pCT and first QACT is not important in this model, it had an average importance of approximately ~40% across all attempts to train the model.

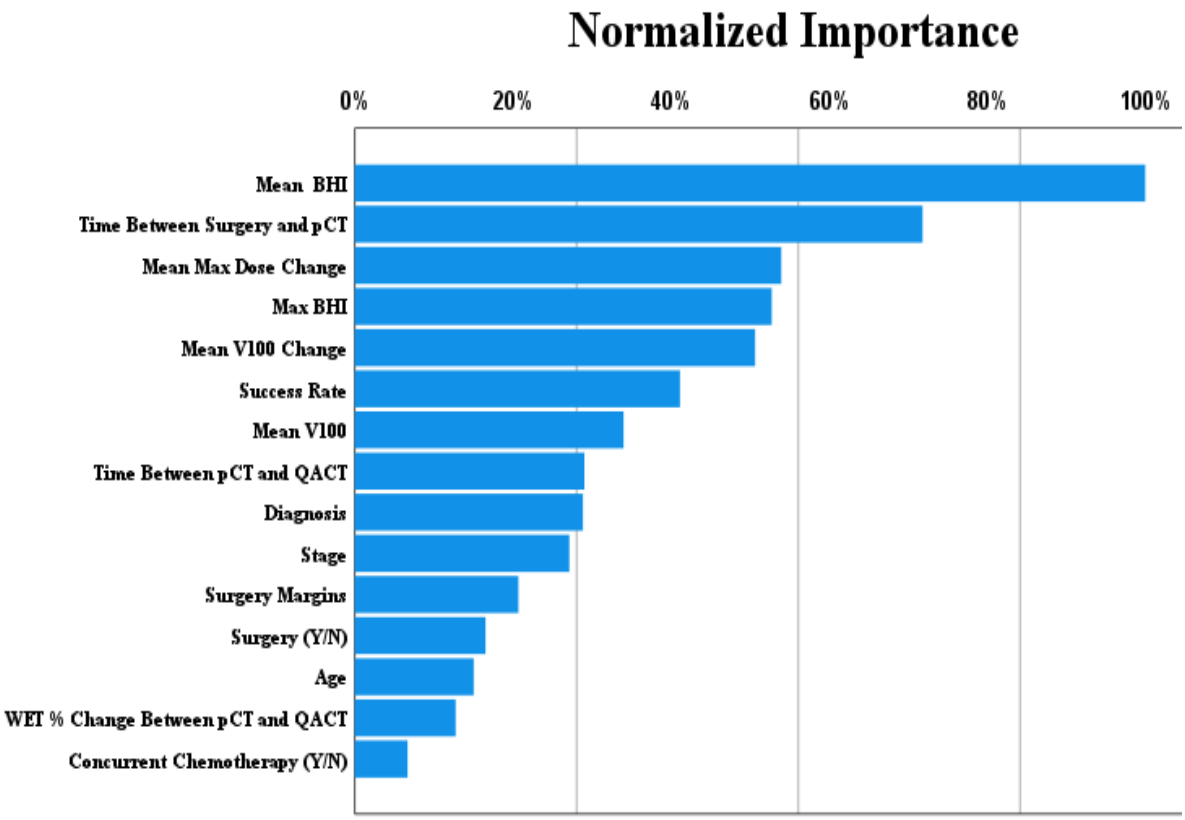

Figure 12 - Normalized importance of features in on-line model. Mean BHI, time between surgery and pCT, and mean max dose change were the most important features.

CHAPTER 4. DISCUSSION

4.1 Clinical Impact of Dosimetric and Clinical Features

Because of proton therapy's sensitivity to anatomical changes, H&N cancers are the most commonly re-planned site at our institution. Reducing re-plan rates would improve efficiency and patient outcomes.

This study analyzed various dosimetric and clinical features between re-plan and non-replan groups to find parameters with statistical significance. Whether or not a feature is statistically significant determines whether it is a good predictor for re-plans.

4.1.1 Significant and Weakly Significant Clinical Features

Tables 3 and 4 show the statistical significance of clinical features. Surgery margins and time between surgery and CT simulation were statistically significant. Age and if patients had surgery are weakly significant. These results agree well with expectations because all these features have a direct impact on anatomical changes, which cause re-plans.

Surgery margins describe the remaining cancer in the edges of the surgery bed. If margins are negative, there are no cancer cells left. If they are positive, not all the cancer has been removed. Close margins are when the cancer cells are close to the edge, but not right on it. These surgery margins impact radiotherapy because the amount of remaining tumor cells will determine whether the tumor bed is treated. This supports the relationship between margins and re-plans.

The time between surgery and pCT simulation has a significant impact on anatomical changes. Typically, missing tissue in the surgery bed is replaced with a tissue flap. Because the body initially rejects the flap, the flap initially swells and then shrinks. Because of this, patients with short times between surgery and pCT have more flap shrinkage occur after simulation than those with long times between surgery and simulation. This leads to patients with shorter time being more likely to be re-planned. This is supported by the data in Table 3, which shows that, on average, patients with re-plans had times that were approximately 2 weeks shorter than patients with no re-plans.

The weak correlation observed between age and re-plans was expected. Surucu et al, which produced a decision tree for predicting tumor shrinkage, found that the patient's age was a good predictor for tumor shrinkage (25). This is supported by the data in Table 3, which shows that, on average, patients with re-plans were approximately 6 years older than patients with no re-plans.

The weak correlation between whether the patient had surgery and re-plans is surprising because, as discussed previously, surgery has an impact on anatomical changes. However, though surgery can impact tissue changes, it can still occur in patients who did not have surgery. Surgery involves the partial or full removal of the tumor. Because a portion of or all the tumor is removed, tumor shrinkage is less of a problem in patients who had surgery. The data in Table 4 supports this, with patients with and without surgery having re-plan rates of 24.2% and 42.3%, respectively.

4.1.2 Non-Significant Clinical Features

As shown in Table 3, sex, stage, T, N, p16, time between pCT and QACT, planner, treatment type, and whether the patient had chemotherapy were not significant.

The sex of the patient does not determine their sensitivity to anatomy changes. Surucu et al found that sex was not a good predictor for tumor shrinkage (25). The data supports this, with male and female patients having similar re-plan rates of 32.5% and 37.5%, respectively.

No patients in our data set had metastatic tumors. This absence of variability led to the lack of correlation between M and re-plans.

Re-plan rates do vary across planners, but this does not make the planner a predictor for re-plans. The differences in re-plan rates between planners are not large enough to expect any correlation between planner and re-plans.

The time between pCT and first QACT is typically 3-4 weeks. This small variability likely is what caused the lack of a correlation with re-plans. This is supported by the data in Table 3, which shows a similar means between re-planned and non-replanned patients of 22 and 23 days, respectively. Another explanation for the lack of significance is that, by the first QACT, most anatomical changes are already detectable. This is supported by the weak significance of the WET change by the first QACT, shown in Table 3. This provides evidence that anatomical changes seen in the first QACT are already good predictors for re-plans, so the time between pCT and QACT is irrelevant.

Only 2 cases out of 132 were palliative. This lack of variation in the type of therapy caused the lack of correlation with re-plans.

The non-significance of site, p16, stage, T, N, and whether the patient had chemotherapy were not expected.

The relationship between site and tumor shrinkage has been observed by Sucuru et al (25). Because p16 varies by site, a correlation between it and re-plans was also expected. A possible explanation for our lack of correlation between site and re-plans is the broad range of sites in H&N cancer, leading to a lack of data for each site and therefore a non-significant result.

Tumor stage, T, and N are expected to be related to re-plans. The stage of the tumor is directly related to the effectiveness of treatment, and thus tumor shrinkage. T and N are related to re-plans similarly. A possible explanation for this lack of correlation in stage, T, and N is the broad range of values they take. In our data, stage, T, and N had 15, 12, and 12 categories, respectively. The individual data in each category is slim and thus a correlation cannot be established.

Because chemotherapy has such a large impact on weight loss, it was expected to be significant. The data showed a large difference between the re-plans rates. Patients with and without chemotherapy had re-plan rates of 39.3% and 26.5%, respectively. This supports the expected result of patients with chemotherapy having higher re-plan rates, but this result is not significant. One possible explanation for this lack of significance is that the type of chemotherapy was not gathered when the clinical data were collected. Certain chemotherapies have greater impact on patient anatomy, and this information could have been obscured. This is supported by the literature (25).

4.1.3 Significant and Weakly Significant Dosimetric Features

The significance of dosimetric features is shown in Table 3. Mean BHI, max BHI, and mean max dose change were all significant. WET change between pCT and the first QACT was weakly significant. These match well with our predictions.

Because the BHI is the ratio of the beam's max dose to its expected contribution, it is a good measure of dose heterogeneity and hotspots. When beams contribute much more dose than expected, anatomical changes impact the plan much more harshly than a less heterogenous plan. Because of this, both mean and max BHI were expected to correlate with re-plans. We found that the more heterogenous a plan is, the more likely it is to be re-planned. This is shown in Table 3. On average, patients with and without re-plans had a mean BHI of 1.11 and 1.09, respectively. A similar relationship is present in max BHI.

Mean max dose change is another measure of hotspots. It describes the shift of the hotspot in response to density changes and anatomy shifts. A high mean max dose change means the plan is not very robust. As expected from less robust plans, patients with high mean max dose changes were more likely to be re-planned. This is supported by the data in Table 3. On average, Patients with and without re-plans had a mean max dose change of 3.22% and 2.87%, respectively.

WET change is the most direct measure of anatomy changes. Because of this, WET change between the pCT and first QACT was expected to correlate with re-plans, but it was barely outside of the confidence interval. With more data, this feature might become significant.

4.1.4 Non-Significant Dosimetric Features

As shown in Table 3, success rate, mean V100, and mean V100 change were not significant. A plausible explanation for the lack of significance is that all 3 of these features must be suitable before a plan is approved for treatment. Approved plans have high success rate and coverage as well as low coverage change, so these features do not differ much between patients. This is supported by Table 3. The mean success rates of re-planned and non re-planned patients are almost equal. Though there is a difference in the means of re-planned and non re-planned patients for mean V100 and mean V100 change, this difference is not significant.

4.2 Clinical Impact of Feature Importance in Neural Networks

As shown in Figures 10 and 12, important features in the pre-treatment and on-line models were mean/max BHI, age, mean max dose change, time between surgery and pCT. These features were all statistically significant.

Mean/max BHI and mean max dose change are the best plan quality parameters to use to predict re-plans. Age and time between surgery and pCT are the best clinical parameters.

4.3 Clinical Impact of Neural Networks for Re-Plan Prediction

The overall prediction accuracies shown in Tables 6 and 7 for the pre-treatment and on-line models are large. The prediction accuracies in the testing data were worse, but not low enough to indicate overfitting. The sensitivity/specificities are high, which is optimal for good prediction models. The ROCs for both models have high area under the curve. In all areas, the on-line prediction model performed better than the pre-treatment model. This was expected because the on-line model included WET changes as an extra feature.

These models are high quality enough to be applied clinically. Using these models could reduce re-plan rates.

4.4 Future Work

In order to be successful, these models need to be implemented in the TPS. A python script could be created that allows the planner to input the necessary dosimetric and clinical features and output the prediction of the neural network based these parameters.

For the on-line model, this script would be used for every QACT. The WET changes between the current QACT and the pCT would be inputted, allowing re-plan predictions to be made for every QACT.

These models can also be refined with better feature selection. Site, stage, and TNM could become significant if more data is gathered or if the categories can be reduced. Conformity index, the ratio of the volume of the CTV covered by the prescription and the volume of the CTV, could also be significant.

CHAPTER 5. CONCLUSION

H&N cancer is a rare, but deadly. Radiotherapy has been used for decades to treat H&N cancer. Recently, proton therapy has been used for treating H&N cancers due to its high dose conformity. Proton therapy is very sensitive to anatomy changes, and H&N patients often have various anatomical changes such as tumor shrinkage during treatment. Because of this, adaptive therapy is necessary for H&N sites, and H&N cancers are the most re-planned site at our institution.

This study looked at various clinical and dosimetric features such as patient surgery and beam heterogeneity and determined how good of predictors they were for re-plans. Using these features, two neural networks for re-plan prediction were made: one before treatment, and one for during treatment.

We several features that were good predictors for re-plans such as average beam heterogeneity, average max dose change, and the time between surgery and pCT. Both models performed well on their training and testing, and both have enough predictive power to be applied clinically.

REFERENCES

[1] Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2018;68:394.

[2] Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer Statistics, 2021. *CA Cancer J Clin.* 2021;71:7.

[3] Emma B. Holliday, Steven J. Frank. Proton Radiation Therapy for Head and Neck Cancer: A Review of the Clinical Experience to Date. *International Journal of Radiation Oncology*. 2014;89:292-302.

[4] Yan D, Vicini F, Wong J, Martinez A. Adaptive radiation therapy. *Phys Med Biol* 1997;42:123–32.

[5] Q. Wu, Y. Chi, P.Y. Chen, D.J. Krauss, D. Yan, A. Martinez. Adaptive replanning strategies accounting for shrinkage in head and neck IMRT. *Int J Radiat Oncol Biol Phys*. 2009;75:924-932

[6] E.K. Hansen, M.K. Bucci, J.M. Quivey, V. Weinberg, P. Xia. Repeat CT imaging and replanning during the course of IMRT for head-and-neck cancer. *Int J Radiat Oncol Biol Phys*. 2006;64:355-62

[7] Müller BS, Duma MN, Kampfer S, Nill S, Oelfke U, Geinitz H, et al. Impact of interfractional changes in head and neck cancer patients on the delivered dose in intensity modulated radiotherapy with protons and photons. *Physica Medica.* 2015; 31: 266–72.

[8] Ahn PH, Lukens JN, Teo B-KK, Kirk M, Lin A. The use of proton therapy in the treatment of head and neck cancers. *The Cancer Journal.* 2014; 20: 421–6.

[9] Szeto YZ, Witte MG, van Kranen SR, Sonke J-J, Belderbos J, van Herk M. Effects of anatomical changes on pencil beam scanning proton plans in locally advanced NSCLC patients. *Radiotherapy and Oncology*. 2016; 120: 286–92.

[10] Chang JY, Li H, Zhu XR, Liao Z, Zhao L, Liu A, et al. Clinical implementation of intensity modulated proton therapy for thoracic malignancies. *Int J Radiat Oncol Biol Phys.* 2014;90:809–18.

[11] Hild S, Graeff C, Rucinski A, Zink K, Habl G, Durante M, et al. Scanned ion beam therapy for prostate carcinoma comparison of single plan treatment and daily plan-adapted treatment. *Strahlenther Onkol.* 2016; 192:118–26.

[12] Knopf A C and Lomax A. In vivo proton range verification: a review. *Phys. Med. Biol.* 2013;58(15):131-60**.**

[13] Grabham P and Sharma P. The effects of radiation on angiogenesis *Vascular Cell*. 2013;5:19.

[14] Schneider U, Pedroni E, Lomax A. The calibration of CT Hounsfield units for radiotherapy treatment planning. *Phys Med Biol*. 1996;41(1):111–24.

[15] Jerry L. Barker, Adam S. Garden, K.Kian Ang, Jennifer C. O'Daniel, He Wang, Laurence E. Court, William H. Morrison, David I. Rosenthal, K.S.Clifford Chao, Susan L. Tucker, Radhe Mohan, Lei Dong. Quantification of volumetric and geometric changes occurring during fractionated radiotherapy for head-and-neck cancer using an integrated CT/linear accelerator system. *International Journal of Radiation Oncology*. 2004; 59(4):960-70

[16] Chencharick, J.D. and Mossman, K.L. (1983), Nutritional consequences of the radiotherapy of head and neck cancer. *Cancer.* 1983;51:811-815.

[17] Donaldson, S.S. and Lenon, R.A. Alterations of nutritional status. Impact of chemotherapy and radiation therapy. *Cancer.* 1979;43:2036-2052.

[18] Cheng Z, Nakatsugawa M, Hu C, et al. Evaluation of classification and regression tree (CART) model in weight loss prediction following head and neck cancer radiation therapy. *Advances in Radiation Oncology*. 2018;3:346–355.

[19] Cheng Z, Nakatsugawa M, Zhou XC, et al. Utility of a Clinical Decision Support System in Weight Loss Prediction After Head and Neck Cancer Radiotherapy. *JCO Clinical Cancer Informatics*. 2019;3:1–11.

[20] Langius JAE, Twisk J, Kampman M, et al. Prediction model to predict critical weight loss in patients with head and neck cancer during (chemo)radiotherapy. *Oral Oncology*. 2016;52:91–96.

[21] White, H., Learning in artificial neural networks: A statistical approach. *Neural Comput.* 1989;1:425–464.

[22] Kajikawa, T., Kadoya, N., Ito, K. et al. Automated prediction of dosimetric eligibility of patients with prostate cancer undergoing intensity-modulated radiation therapy using a convolutional neural network. *Radiol Phys Technol.* 2018;11**:**320–327.

[23] Isaksson, M., Jalden, J. and Murphy, M.J. On using an adaptive neural network to predict lung tumor motion during respiration for radiotherapy applications. *Med. Phys.* 2005;32:3801-3809.

[24] Song, Y., Hu, J., Liu, Y. Hu, H., Huang, Y., Bai, S.,and Yi, Z. Dose prediction using a deep neural network for accelerated planning of rectal cancer radiotherapy. *Radiotherapy and Oncology*. 2020;149:11-116.

[25] Surucu M, Shah KK, Mescioglu I, et al. Decision Trees Predicting Tumor Shrinkage for Head and Neck Cancer: Implications for Adaptive Radiotherapy. *Technology in Cancer Research & Treatment*. 2016:15:139-145.